

# Optimal Nonstationary Reproduction Distribution for Nonanticipative RDF on Abstract Alphabets

Photios A. Stavrou, Charalambos D. Charalambous and Christos K. Kourtellaris

ECE Department, University of Cyprus, Nicosia, Cyprus

Email: {stavrou.fotios, chadcha, kourtellaris.christos}@ucy.ac.cy

**Abstract**—THIS PAPER IS ELIGIBLE FOR THE STUDENT PAPER AWARD. In this paper we introduce a definition for nonanticipative Rate Distortion Function (RDF) on abstract alphabets, and we invoke weak convergence of probability measures to show various of its properties, such as, existence of the optimal reproduction conditional distribution, compactness of the fidelity set, lower semicontinuity of the RDF functional, etc. Further, we derive the closed form expression of the optimal nonstationary reproduction distribution. This expression is computed recursively backward in time. Throughout the paper we point out an operational meaning of the nonanticipative RDF by recalling the coding theorem derive in [1], and we state relations to Gorbunov-Pinsker's nonanticipatory  $\epsilon$ -entropy [2].

## I. INTRODUCTION

Motivated by communication and control applications in which lossy nonanticipative reproductions of nonstationary sources subject to a fidelity set are desirable, in this paper we introduce a definition for nonanticipative RDF.

One envisioned application is quantization of two dimensional sources, represented by space-time processes, in which an IID assumption is imposed on the spatial index, when the time index is fixed. Such a model is considered in [1] to derive a coding theorem which states that the Optimal Performance Theoretically Achievable (OPTA) by sequential quantizers is given by the so-called sequential RDF, expressed in terms of mutual information and a conditional independence imposed on the fidelity set. Another envisioned application is source-channel matching [3], [4], in which nonanticipation of the reproduction distribution is a necessary conditions for the realization of this conditional distribution by encoder-channel-decoder maps that operate without anticipation, and hence the delay incurred on the end-to-end system is limited, or uncoded transmission [3].

In this paper, the nonanticipative RDF is defined on abstract alphabets, using an information measure which is a special case of directed information [5] from the source sequence to the reproduction sequence. The main contributions are the treatment of the nonanticipative RDF without assuming stationarity of the source, the derivation of various results regarding existence of the optimal nonstationary reproduction distribution, its closed form expression via backward recursions, relations to existing coding theorems, and relations to the nonanticipatory  $\epsilon$ -entropy introduced by Gorbunov and Pinsker [2]. Since nonstationary nonanticipative RDF embeds the stationary nonanticipative RDF, under certain assumptions,

our results apply for the stationary case too. Note that our results extend the bounds developed recently in [6] for stationary Gaussian sources to general sources.

This paper is organized as follows. Section II formulates the nonanticipative RDF on abstract spaces. Section III, discusses the relation between nonanticipative RDF with Gorbunov-Pinsker's nonanticipatory  $\epsilon$ -entropy and sequential RDF. Section IV provides the conditions under which the existence of the optimal reproduction conditional distribution of nonanticipative RDF is derived. Finally, Section V gives the optimal solution for nonstationary processes. Throughout the paper we also include comments on the reduction of our results to the stationary nonanticipative RDF.

## II. NONANTICIPATIVE RDF ON ABSTRACT SPACES

In this section, we define the information theoretic nonanticipative RDF based on the methodology described in [7].

**Notation.** Let  $\mathbb{N} \triangleq \{0, 1, 2, \dots\}$ , and  $\mathbb{N}^n \triangleq \{0, 1, 2, \dots, n\}$ . Introduce two sequence of spaces  $\{(\mathcal{X}_n, \mathcal{B}(\mathcal{X}_n)) : n \in \mathbb{N}\}$  and  $\{(\mathcal{Y}_n, \mathcal{B}(\mathcal{Y}_n)) : n \in \mathbb{N}\}$ , where  $\mathcal{X}_n, \mathcal{Y}_n, n \in \mathbb{N}$ , are Polish spaces, and  $\mathcal{B}(\mathcal{X}_n)$  and  $\mathcal{B}(\mathcal{Y}_n)$  are Borel  $\sigma$ -algebras of subsets of  $\mathcal{X}_n$  and  $\mathcal{Y}_n$ , respectively. Points in  $\mathcal{X}^{\mathbb{N}} \triangleq \times_{n \in \mathbb{N}} \mathcal{X}_n$ ,  $\mathcal{Y}^{\mathbb{N}} \triangleq \times_{n \in \mathbb{N}} \mathcal{Y}_n$  are denoted by  $\mathbf{x} \triangleq \{x_0, x_1, \dots\} \in \mathcal{X}^{\mathbb{N}}$ ,  $\mathbf{y} \triangleq \{y_0, y_1, \dots\} \in \mathcal{Y}^{\mathbb{N}}$ , respectively, while their restrictions to finite coordinates by  $x^n \triangleq \{x_0, x_1, \dots, x_n\} \in \mathcal{X}_{0,n}$ ,  $y^n \triangleq \{y_0, y_1, \dots, y_n\} \in \mathcal{Y}_{0,n}$ , for  $n \in \mathbb{N}$ .

Let  $\mathcal{B}(\mathcal{X}^{\mathbb{N}}) \triangleq \odot_{i \in \mathbb{N}} \mathcal{B}(\mathcal{X}_i)$  denote the  $\sigma$ -algebra on  $\mathcal{X}^{\mathbb{N}}$  generated by cylinder sets and similarly for  $\mathcal{B}(\mathcal{Y}^{\mathbb{N}}) \triangleq \odot_{i \in \mathbb{N}} \mathcal{B}(\mathcal{Y}_i)$ , while  $\mathcal{B}(\mathcal{X}_{0,n})$  and  $\mathcal{B}(\mathcal{Y}_{0,n})$  denote the  $\sigma$ -algebras with bases over  $A_i \in \mathcal{B}(\mathcal{X}_i)$ , and  $B_i \in \mathcal{B}(\mathcal{Y}_i)$ ,  $i = 0, 1, \dots, n$ , respectively. Let  $\mathcal{Q}(\mathcal{Y}; \mathcal{X})$  denote the set of stochastic kernels on  $\mathcal{Y}$  given  $\mathcal{X}$  [8].

### A. Construction of the Measures

**Source Distribution.** The source distribution  $\{p_n(dx_n; x^{n-1}) : n \in \mathbb{N}\}$  satisfies the following conditions.

- i)** For  $n \in \mathbb{N}$ ,  $p_n(\cdot; x^{n-1})$  is a probability measure on  $\mathcal{B}(\mathcal{X}_n)$ ;
- ii)** For every  $A_n \in \mathcal{B}(\mathcal{X}_n)$ ,  $n \in \mathbb{N}$ ,  $p_n(A_n; x^{n-1})$  is a  $\odot_{i=0}^{n-1} \mathcal{B}(\mathcal{X}_i)$ -measurable function of  $x^{n-1} \in \mathcal{X}_{0,n-1}$ .

Any distribution satisfying **i)**, **ii)** is denoted by  $p_n(\cdot; \cdot) \in \mathcal{Q}(\mathcal{X}_n; \mathcal{X}_{0,n-1})$ .

Let  $A \in \mathcal{B}(\mathcal{X}_{0,n})$  be a cylinder set of the form  $A \triangleq \{\mathbf{x} \in \mathcal{X}^{\mathbb{N}} : x_0 \in A_0, x_1 \in A_1, \dots, x_n \in A_n\}$ ,  $A_i \in$

$\mathcal{B}(\mathcal{X}_i)$ ,  $i = 0, 1, \dots, n$ . Define a family of measures  $\mathbf{P}(\cdot)$  on  $\mathcal{B}(\mathcal{X}^{\mathbb{N}})$ , denoted by the set  $\mathcal{M}_1(\mathcal{X}^{\mathbb{N}})$ , given by

$$\mathbf{P}(A) \triangleq \int_{A_0} p_0(dx_0) \dots \int_{A_n} p_n(dx_n; x^{n-1}) \equiv \mu_{0,n}(A_{0,n}) \quad (1)$$

where  $A_{0,n} = \times_{i=0}^n A_i$ . The notation  $\mu_{0,n}(\cdot)$  is used to denote the restriction of the measure  $\mathbf{P}(\cdot)$  on cylinder sets  $A \in \mathcal{B}(\mathcal{X}_{0,n})$ , for  $n \in \mathbb{N}$ .

**Reproduction Distribution.** The reproduction distribution  $\{q_n(dy_n; y^{n-1}, x^n) : n \in \mathbb{N}\}$  satisfies the following conditions.

**iv)** For  $n \in \mathbb{N}$ ,  $q_n(\cdot | y^{n-1}, x^n)$  is a probability measure on  $\mathcal{B}(\mathcal{Y}_n)$ ;

**v)** For every  $B_n \in \mathcal{B}(\mathcal{Y}_n)$ ,  $n \in \mathbb{N}$ ,  $q_n(B_n; y^{n-1}, x^n)$  is a  $\odot_{i=0}^{n-1}(\mathcal{B}(\mathcal{Y}_i) \odot \mathcal{B}(\mathcal{X}_i)) \odot \mathcal{B}(\mathcal{X}_n)$ -measurable function of  $x^n \in \mathcal{X}_{0,n}$ ,  $y^{n-1} \in \mathcal{Y}_{0,n-1}$ .

Given a cylinder set  $B \triangleq \{\mathbf{y} \in \mathcal{Y}^{\mathbb{N}} : y_0 \in B_0, y_1 \in B_1, \dots, y_n \in B_n\}$ , define a family of measures  $\mathbf{Q}(\cdot | \mathbf{x})$  on  $\mathcal{B}(\mathcal{Y}^{\mathbb{N}})$  by

$$\mathbf{Q}(B | \mathbf{x}) \triangleq \int_{B_0} q_0(dy_0; x_0) \dots \int_{B_n} q_n(dy_n; y^{n-1}, x^n) \quad (2)$$

$$\equiv \vec{\mathcal{Q}}_{0,n}(B_{0,n} | x^n), \quad B_{0,n} \in \mathcal{B}(\mathcal{Y}_{0,n}). \quad (3)$$

Then  $\mathbf{Q}(\cdot | \mathbf{x})$  is a unique measure on  $(\mathcal{Y}^{\mathbb{N}}, \mathcal{B}(\mathcal{Y}^{\mathbb{N}}))$  for which the family of distributions  $\{q_n(dy_n; y^{n-1}, x^n) : n \in \mathbb{N}\}$  is obtained.

Consider any family of measures  $\mathbf{Q}(\cdot | \mathbf{x})$  on  $\mathcal{B}(\mathcal{Y}^{\mathbb{N}})$  satisfying the following consistency condition.

**C1:** If  $D \in \mathcal{B}(\mathcal{Y}_{0,n})$ , then  $\mathbf{Q}(D | \mathbf{x})$  is  $\mathcal{B}(\mathcal{X}_{0,n})$ -measurable function of  $\mathbf{x} \in \mathcal{X}^{\mathbb{N}}$ .

The set of such measures is denoted by  $\mathcal{Q}^{\mathbf{C1}}(\mathcal{Y}^{\mathbb{N}}, \mathcal{X}^{\mathbb{N}})$ .

Then, for any family of measures  $\mathbf{Q}(\cdot | \mathbf{x})$  on  $(\mathcal{Y}^{\mathbb{N}}, \mathcal{B}(\mathcal{Y}^{\mathbb{N}}))$  satisfying consistency condition **C1** one can construct a collection of functions  $\{q_n(dy_n; y^{n-1}, x^n) : n \in \mathbb{N}\}$  satisfying conditions **iv)** and **v)** which are connected with  $\mathbf{Q}(\cdot | \mathbf{x})$  via relation (2) (see also [7]).

By following the methodology in [7], given the basic measures  $\mathbf{P}(\cdot)$  on  $\mathcal{X}^{\mathbb{N}}$  and  $\mathbf{Q}(\cdot | \mathbf{x})$  on  $\mathcal{Y}^{\mathbb{N}}$  satisfying consistency condition **C1**, we can uniquely define the collection of conditional distributions  $\{p_n(\cdot; \cdot) \in \mathcal{Q}(\mathcal{X}_n; \mathcal{X}_{0,n-1}) : n \in \mathbb{N}\}$  via (1), and  $\{q_n(\cdot; \cdot, \cdot) \in \mathcal{Q}(\mathcal{Y}_n; \mathcal{Y}_{0,n-1} \times \mathcal{X}_{0,n}) : n \in \mathbb{N}\}$  via (2), respectively, and vice versa, hence the distribution of the RV's  $\{(X_i, Y_i) : i \in \mathbb{N}\}$  is well defined.

Next, we introduce the information definition of nonanticipative RDF. Given the source distribution  $\mathbf{P}(\cdot) \in \mathcal{M}_1(\mathcal{X}^{\mathbb{N}})$  and reproduction distribution  $\mathbf{Q}(\cdot | \cdot) \in \mathcal{Q}^{\mathbf{C1}}(\mathcal{Y}^{\mathbb{N}}; \mathcal{X}^{\mathbb{N}})$  define the following measures.

**P1:** The joint distribution on  $\mathcal{X}^{\mathbb{N}} \times \mathcal{Y}^{\mathbb{N}}$  defined uniquely by

$$(\mu_{0,n} \otimes \vec{\mathcal{Q}}_{0,n})(\times_{i=0}^n A_i \times B_i), \quad A_i \in \mathcal{B}(\mathcal{X}_i), \quad B_i \in \mathcal{B}(\mathcal{Y}_i).$$

**P2:** The marginal distribution on  $\mathcal{Y}^{\mathbb{N}}$  defined uniquely for  $B_i \in \mathcal{B}(\mathcal{Y}_i)$ ,  $i = 0, 1, \dots, n$ , by

$$\nu_{0,n}(\times_{i=0}^n B_i) = (\vec{\mathcal{P}}_{0,n} \otimes \vec{\mathcal{Q}}_{0,n})(\times_{i=0}^n (\mathcal{X}_i \times B_i)).$$

**P3:** The product distribution  $\vec{\Pi}_{0,n} : \mathcal{B}(\mathcal{X}_{0,n}) \odot \mathcal{B}(\mathcal{Y}_{0,n}) \mapsto [0, 1]$  defined uniquely for  $A_i \in \mathcal{B}(\mathcal{X}_i)$ ,  $B_i \in \mathcal{B}(\mathcal{Y}_i)$ ,  $i = 0, 1, \dots, n$ , by

$$\vec{\Pi}_{0,n}(\times_{i=0}^n (A_i \times B_i)) \triangleq (\mu_{0,n} \otimes \nu_{0,n})(\times_{i=0}^n (A_i \times B_i)).$$

The information theoretic measure of interest is a special case of directed information [1] defined by relative entropy  $\mathbb{D}(\cdot || \cdot)^1$

$$I_{\mu_{0,n}}(X^n \rightarrow Y^n) \triangleq \mathbb{D}(\mu_{0,n} \otimes \vec{\mathcal{Q}}_{0,n} || \vec{\Pi}_{0,n}) \quad (4)$$

$$= \int \log \left( \frac{\vec{\mathcal{Q}}_{0,n}(dy^n | x^n)}{\nu_{0,n}(dy^n)} \right) (\mu_{0,n} \otimes \vec{\mathcal{Q}}_{0,n})(dx^n, dy^n) \quad (5)$$

$$\equiv \mathbb{I}_{X^n \rightarrow Y^n}(\mu_{0,n}, \vec{\mathcal{Q}}_{0,n}). \quad (6)$$

The equivalence between (4) and (5) follows from the Radon-Nikodym Derivative (RND). The notation  $\mathbb{I}_{X^n \rightarrow Y^n}(\cdot, \cdot)$  indicates the functional dependence of  $I_{\mu_{0,n}}(X^n \rightarrow Y^n)$  on  $\{\mu_{0,n}, \vec{\mathcal{Q}}_{0,n}\}$ .

### B. Nonanticipative RDF

We are now ready to introduce the information definition of nonanticipative RDF. The distortion function  $d_{0,n}(x^n, y^n) : \mathcal{X}_{0,n} \times \mathcal{Y}_{0,n} \mapsto [0, \infty)$  is a measurable function, and the fidelity of reproduction is defined by

$$\begin{aligned} \vec{\mathcal{Q}}_{0,n}(D) &\triangleq \left\{ \vec{\mathcal{Q}}_{0,n} \in \mathcal{Q}^{\mathbf{C1}}(\mathcal{Y}_{0,n}; \mathcal{X}_{0,n}) : \right. \\ \ell_{d_{0,n}}(\vec{\mathcal{Q}}_{0,n}) &\triangleq \frac{1}{n+1} \int d_{0,n}(x^n, y^n) (\mu_{0,n} \otimes \vec{\mathcal{Q}}_{0,n})(dx^n, dy^n) \\ &\left. \leq D \right\}, \quad D \geq 0. \end{aligned} \quad (7)$$

The information nonanticipative RDF is defined by

$$\vec{R}_{0,n}(D) \triangleq \inf_{\vec{\mathcal{Q}}_{0,n} \in \vec{\mathcal{Q}}_{0,n}(D)} \mathbb{I}_{X^n \rightarrow Y^n}(\mu_{0,n}, \vec{\mathcal{Q}}_{0,n}). \quad (8)$$

If the infimum in (8) does not exist then  $\vec{R}_{0,n}(D) = \infty$ .

The nonanticipative RDF rate is defined by

$$\vec{R}(D) = \lim_{n \rightarrow \infty} \frac{1}{n+1} \vec{R}_{0,n}(D) \quad (9)$$

provided the limit exists.

### III. NONANTICIPATORY $\epsilon$ -ENTROPY, MESSAGE GENERATION RATES AND SEQUENTIAL RDF

In this section, we establish some preliminary relations between (8), (9) and 1) Gorbunov-Pinsker's nonanticipatory  $\epsilon$ -entropy and message generation rates [2] and 2) sequential RDF and coding theorem [1].

<sup>1</sup>Unless stated otherwise, integrals with respect to measures are over the spaces on which these are defined.

### A. Nonanticipatory $\epsilon$ -Entropy and Message Generation Rates

We recall Gorbunov-Pinsker's definition of nonanticipatory  $\epsilon$ -entropy [2]. Given a source  $P_{X^n} \in \mathcal{M}_1(\mathcal{X}_{0,n})$  and a reproduction  $P_{Y^n|X^n} \in \mathcal{Q}(\mathcal{Y}_{0,n}; \mathcal{X}_{0,n})$  introduce the fidelity set

$$Q_{0,n}(D) \triangleq \left\{ P_{Y^n|X^n}(dy^n|x^n) : \frac{1}{n+1} \int d_{0,n}(x^n, y^n) P_{Y^n|X^n}(dy^n|x^n) \otimes P_{X^n}(dx^n) \leq D \right\}.$$

Next we introduce two definitions from [2].

**Definition 1. (1)** The source  $X^\infty \triangleq \{X_i : i \in \mathbb{N}\}$  is called "specified" if  $P_{Y^k|X^k} \in Q_{0,k}(D)$  and  $P_{Y_{k+1}^n|X_{k+1}^n} \in Q_{k+1,n}(D)$  implies  $P_{Y^k, Y_{k+1}^n|X^k, X_{k+1}^n} \in Q_{0,n}(D)$ , and it is called "consistent" if the reverse holds.

**(2)** The source is called stationary if  $\{X_i : i \in \mathbb{N}\}$  is stationary and for any  $k$ ,  $Q_{0,n}(D)$  and  $Q_{k,n+k}(D)$  are copies of the same set.

Typical example are stationary sources with single letter fidelities  $d_{0,n}(x^n, y^n) = \sum_{i=0}^n \rho(x_i, y_i)$ . For a specified source, Gorbunov and Pinsker restricted the set  $Q_{0,n}(D)$  to those reproduction distributions which satisfy the Markov chain (MC)  $X_{n+1}^\infty \leftrightarrow X^n \leftrightarrow Y^n \Leftrightarrow P_{Y^n|X^\infty}(dy^n|x^\infty) = P_{Y^n|X^n}(dy^n|x^n) - a.s., n = 0, 1, \dots$ , and introduced the nonanticipatory  $\epsilon$ -entropy defined by

$$R_{0,n}^{na}(D) \triangleq \inf_{\substack{P_{Y^n|X^n} \in Q_{0,n}(D): \\ X_{i+1}^n \leftrightarrow X^i \leftrightarrow Y^i, i=0,1,\dots,n-1}} I(X^n; Y^n). \quad (10)$$

If the infimum in (10) does not exist then  $R_{0,n}^{na}(D) = \infty$ .

The difference between the classical RDF [9] and (10) is the presence of the MC, which implies that for each  $i$ ,  $Y_i$  is a function of the past and present source symbols  $X^i$ , and it is independent of the future source symbols  $X_{i+1}^n$ ,  $n \in \mathbb{N}$ .

Moreover, the authors in [2] also introduced the nonanticipatory message generation rate of the source defined by

$$R^{na}(D) = \lim_{n \rightarrow \infty} \frac{1}{n+1} R_{0,n}^{na}(D) \quad (11)$$

provided the limit exists.

An alternative definition of the nonanticipatory message generation rate of the source is defined by [2]

$$R^{na,+}(D) = \inf_{\substack{P_{Y^\infty|X^\infty} \in Q_{0,\infty}(D): \\ X_{i+1}^\infty \leftrightarrow X^i \leftrightarrow Y^i, i=0,1,\dots}} \lim_{n \rightarrow \infty} \frac{I(X^n; Y^n)}{n+1} \quad (12)$$

whenever the limit exists. Clearly, in general  $R^{na,+}(D) \geq R^{na}(D)$ .

The main results derived in [2] are the following.

**GP 1:** If the source is stationary for some finite  $k$ , i.e.,  $R_{0,k}^{na}(D) < d = \text{constant} < \infty$ , then  $R^{na}(D)$  is defined and it is finite, and  $R^{na}(D) \geq R^{na,+}(D)$ , hence  $R^{na}(D) = R^{na,+}(D)$ .

**GP 2:** If the source is stationary and consistent then  $R^{na}(D)$  is always defined, and either 1) for some finite  $k$ ,  $R_{0,k}^{na}(D) < \infty$  which implies

$\lim_{k \rightarrow \infty} \frac{1}{k+1} R_{0,k}^{na}(D) = R^{na}(D) < \infty$ , or 2) for any  $k \in \mathbb{N}$ ,  $R_{0,k}^{na}(D) = R^{na}(D) = \infty$ .

**GP 3:** For a single letter distortion and stationary source, if  $R^{na}(D) = R^{na,+}(D)$  then the analysis of infimum in (12) is realizable in terms of stationary source-reproduction pairs  $\{(X_i, Y_i) : i \in \mathbb{N}\}$  such that  $P_{Y^n|X^\infty}(dy^n|x^\infty) = P_{Y^n|X^n}(dy^n|x^n)$  and  $P_{Y^n|X^n} \otimes P_{X^n} \in Q_{0,n}(D)$ .

With respect to Gorbunov-Pinsker's definition of nonanticipatory  $\epsilon$ -entropy and message generation rate, 1) we show that (10) reduces to the information definition of nonanticipative RDF (8), 2) under certain general conditions the infimum in (10) exists and it is finite, and, 3) we derive an expression of the optimal reproduction distribution which achieves the infimum in (8) or (10).

First, we establish the connection between nonanticipatory  $\epsilon$ -entropy (10) and nonanticipative RDF (8) by utilizing the following general equivalent statements of MCs.

**Lemma 1.** The following are equivalent for each  $n \in \mathbb{N}$ .

- (1)  $P_{Y^n|X^n}(dy^n|x^n) = \vec{P}_{Y^n|X^n}(dy^n|x^n)$ -a.s.;
- (2) For each  $i = 0, 1, \dots, n-1$ ,  $Y_i \leftrightarrow (X^i, Y^{i-1}) \leftrightarrow (X_{i+1}, X_{i+2}, \dots, X_n)$ , forms a MC;
- (3) For each  $i = 0, 1, \dots, n-1$ ,  $Y^i \leftrightarrow X^i \leftrightarrow X_{i+1}$  forms a MC;
- (4) For each  $i = 0, 1, \dots, n-1$ ,  $X_{i+1}^n \leftrightarrow X^i \leftrightarrow Y^i$ , forms a MC.

*Proof:* The derivation is straightforward. ■

By utilizing Lemma 1, then  $\vec{R}_{0,n}(D) = R_{0,n}^{na}(D)$ . Then the extremum of the nonanticipatory  $\epsilon$ -entropy (10) is equivalent to the extremum of  $\vec{R}_{0,n}(D)$  given by (8).

### B. Sequential RDF and Coding Theorems

Next, we establish a coding theorem for  $\vec{R}_{0,n}(D)$  using the information sequential RDF introduced by Tatikonda in [1], which utilizes a similar formulation to the nonanticipatory  $\epsilon$ -entropy. The coding theorem is derived by considering a two dimensional source  $X^{n,s} \triangleq \{X_{i,j} : i = 0, \dots, n, j = 0, \dots, s\} \in \otimes_{i=0}^n \otimes_{j=0}^s \mathcal{X}_{i,j}$ , where  $i$  represents time index and  $j$  represents spatial index. The coding theorem is based on the following definitions.

**Definition 2.** A sequential quantizer is a sequence of measurable functions  $f^n = \{f_i : i = 0, 1, \dots, n\}$  defined by  $f_i : \mathcal{X}^{i,s} \times \mathcal{Y}^{i-1,s} \mapsto \mathcal{Y}_i^s$ ,  $\mathcal{Y}_i^s = f_i(x^{i,s}, y^{i-1,s})$ ,  $i = 1, \dots, n$ . The set of all such quantizers is denoted by  $\mathcal{F}^{n,s}$ .

**Definition 3.** Let  $Q_{0,n,s}^{SRD,o}(D)$  denote the fidelity set

$$Q_{0,n,s}^{SRD,o} \triangleq \left\{ f^n \in \mathcal{F}^{n,s} : \frac{1}{n+1} \sum_{i=0}^n \mathbb{E}_{P_{X^{i,s}}} \rho_s(X_i^s, Y_i^s) \leq D \right\}$$

where  $\rho_s : \mathcal{X}_i^s \times \mathcal{Y}_i^s \mapsto [0, \infty) : i = 0, 1, \dots, n$  is measurable. The operational sequential RDF is defined by

$$R_{0,n,s}^{SRD,o}(D) \triangleq \inf_{f^n \in Q_{0,n,s}^{SRD,o}(D)} \frac{1}{(n+1)(s+1)} H(Y_0^s, \dots, Y_n^s)$$

and the operational sequential RDF rate is defined by

$$R^{SRD,o}(D) \triangleq \lim_{s \rightarrow \infty} R_{0,n,s}^{SRD,o}(D).$$

The information sequential RDF for which a coding theorem is derived in [1] is the following. Given the two dimensional source  $P_{X^{n,s}}(dx^{n,s})$ , a reproduction distribution  $P_{Y^{n,s}|X^{n,s}}(dy^{n,s}|x^{n,s})$ , and a fidelity set

$$Q_{0,n,s}^{SRD}(D) \triangleq \left\{ P_{Y^{n,s}|X^{n,s}}(dy^{n,s}|x^{n,s}) : \frac{1}{n+1} \sum_{i=0}^n \mathbb{E}_{P_{X_i^s, Y_i^s}} \rho_s(X_i^s, Y_i^s) \leq D \right\} \quad (13)$$

the information sequential RDF is defined by

$$R_{0,n,s}^{SRD}(D) = \inf_{\substack{P_{Y^{n,s}|X^{n,s}} \in Q_{0,n,s}^{SRD}(D) \\ (X_{i+1}^s, \dots, X_n^s) \leftrightarrow (X_i^s, Y^{i-1,s}) \leftrightarrow Y_i^s, i=0,1,\dots,n}} \left\{ \frac{1}{(n+1)(s+1)} I(X^{n,s}; Y^{n,s}) \right\}. \quad (14)$$

The sequential source coding theorem is the following.

**Theorem 1.** (Sequential Source Coding Theorem [1]) Suppose  $\{X_{i,j} : i = 0, 1, \dots, n, j = 0, 1, \dots, s\}$  are finite alphabets,  $P_{X^{n,s}}(dx^{n,s}) = \otimes_{j=0}^s P(dx_j^n)$ ,  $\{X_j^n : j = 0, 1, \dots, s\}$  identically distributed, and there exists an  $x_0$  and  $D_{max} > 0$  such that  $\mathbb{E}_{P_{X_{i,j}}} \rho_s(X_{i,j}, x_0) < D_{max}$ , for all  $i = 0, 1, \dots, n$ ,  $j = 0, 1, \dots, s$ . Then for any  $\epsilon > 0$  and finite  $n \in \mathbb{N}$ , there exists  $s(\epsilon, n)$  such that for all  $s \geq s(\epsilon, n)$  we have

$$R_{0,n,s}^{SRD,o}(D + \epsilon) \leq R_{0,n}^{SRD}(D) + \epsilon$$

where

$$R_{0,n}^{SRD}(D) \triangleq \inf_{\substack{P_{Y^n|X^n} : \frac{1}{n+1} \mathbb{E}_{P_{X^n, Y^n}} \{\sum_{i=0}^n \rho(X_i, Y_i) \leq D\} \\ X_{i+1}^n \leftrightarrow (X_i^i, Y^{i-1}) \leftrightarrow Y_i^i, i=0,1,\dots,n}} \frac{I(X^n; Y^n)}{n+1}. \quad (15)$$

Notice that  $R_{0,n}^{SRD}(D)$  is precisely Gorbunov-Pinsker's nonanticipatory  $\epsilon$ -entropy  $\frac{1}{n+1} R_{0,n}^{na}(D)$  [2]. Moreover, by Lemma 1  $R_{0,n}^{SRD}(D) \equiv \frac{1}{n+1} \bar{R}_{0,n}(D)$  given by (8). However, the coding Theorem 1 is valid for finite time index  $n$ , since its derivation is based on taking  $s \rightarrow \infty$ . Consequently, with respect to the information sequential RDF (15), our objective is to find the expression of the nonstationary optimal reproduction distribution.

#### IV. EXISTENCE OF REPRODUCTION CONDITIONAL DISTRIBUTION OF NONANTICIPATIVE RDF

In this section, the existence of the minimizing  $(n+1)$ -fold convolution of conditional distributions in (8) is established by using the topology of weak convergence of probability measures on Polish spaces. In fact, our results are more general than what is envisioned by the assumptions in [1], [2], since we work with abstract Polish spaces and general distortion functions. First, we recall some properties from [7].

**Theorem 2.** [7] Let  $\{\mathcal{X}_n : n \in \mathbb{N}\}$  and  $\{\mathcal{Y}_n : n \in \mathbb{N}\}$  be Polish spaces. Then

- (1) The set  $\mathcal{Q}^{C1}(\mathcal{Y}_{0,n}; \mathcal{X}_{0,n})$  is convex.
- (2)  $\mathbb{I}_{X^n \rightarrow Y^n}(\mu_{0,n}, \bar{Q}_{0,n})$  is a convex functional of  $\bar{Q}_{0,n} \in \mathcal{Q}^{C1}(\mathcal{Y}_{0,n}; \mathcal{X}_{0,n})$  for a fixed  $\mu_{0,n} \in \mathcal{M}_1(\mathcal{X}_{0,n})$ .
- (3) The set  $\bar{Q}_{0,n}(D)$  is convex.

Let  $BC(\mathcal{Y}_{0,n})$  denotes the set of bounded continuous real-valued functions on  $\mathcal{Y}_{0,n}$ . Below, we introduce the main conditions for the existence of nonanticipative RDF (8).

**Assumption 1.** The following are assumed.

- (A1)  $\mathcal{Y}_{0,n}$  is a compact Polish space,  $\mathcal{X}_{0,n}$  is a Polish space;
- (A2) For all  $h(\cdot) \in BC(\mathcal{Y}_{0,n})$ , the function mapping  $(x^n, y^{n-1}) \in \mathcal{X}_{0,n} \times \mathcal{Y}_{0,n-1} \mapsto \int_{\mathcal{Y}_n} h(y) q_n(dy; y^{n-1}, x^n) \in \mathbb{R}$  is continuous jointly in the variables  $(x^n, y^{n-1}) \in \mathcal{X}_{0,n} \times \mathcal{Y}_{0,n-1}$ ;
- (A3)  $d_{0,n}(x^n, \cdot) : \mathcal{Y}_{0,n} \mapsto [0, \infty)$  is continuous on  $\mathcal{Y}_{0,n}$ , uniformly in  $x^n \in \mathcal{X}_{0,n}$ ;
- (A4) There exist sequence  $(x^n, y^n) \in \mathcal{X}_{0,n} \times \mathcal{Y}_{0,n}$  satisfying  $d_{0,n}(x^n, y^n) < D$ .

The following weak compactness result can be obtained, which will be used to establish existence.

**Lemma 2.** [4] Suppose Assumption 1 (A1), (A2) hold. Then

- (1) The set  $\mathcal{Q}^{C1}(\mathcal{Y}_{0,n}; \mathcal{X}_{0,n})$  is weakly compact.
- (2) Under the additional conditions (A3), (A4) the set  $\bar{Q}_{0,n}(D)$  is a closed subset of  $\mathcal{Q}^{C1}(\mathcal{Y}_{0,n}; \mathcal{X}_{0,n})$  (hence compact).

The next theorem establishes existence of the minimizing reproduction distribution for (8). First, we need the following Lemma.

**Lemma 3.** [7] Under Assumptions 1 (A1), (A2),  $\mathbb{I}_{X^n \rightarrow Y^n}(\mu_{0,n}, \bar{Q}_{0,n})$  is lower semicontinuous on  $\bar{Q}_{0,n} \in \mathcal{Q}^{C1}(\mathcal{Y}_{0,n}; \mathcal{X}_{0,n})$  for a fixed  $\mu_{0,n} \in \mathcal{M}_1(\mathcal{X}_{0,n})$ .

Next, we state the main Theorem.

**Theorem 3.** (Existence [4]) Suppose Assumption 1 hold. Then  $\bar{R}_{0,n}(D)$  has a minimum.

Thus, Theorem 3 implies the following results. By **GP1**, for a stationary source  $R^{na}(D) = R^{na,+}(D)$  is defined and it is finite, by **GP2**, for a stationary consistent source  $\lim_{k \rightarrow \infty} \frac{1}{k+1} R_{0,k}^{na}(D) = R^{na}(D) < \infty$ , and by **GP3**, for a stationary source and single letter distortion,  $R^{na}(D) = R^{na,+}(D)$  and the infimum in (12) is realized by stationary source-reproduction pairs.

#### V. OPTIMAL REPRODUCTION OF NONANTICIPATIVE RDF

In this section, we derive the expression of reproduction conditional distribution which achieves the infimum of  $\bar{R}_{0,n}(D)$  or  $R_{0,n}^{SRD}(D)$ . We assume a distortion function of the form  $d_{0,n}(x^n, y^n) \triangleq \sum_{i=0}^n \rho_{0,i}(x^i, y^i)$ . We shall need, the Gateaux

differential of  $\mathbb{I}_{X^n \rightarrow Y^n}(\mu_{0,n}, \vec{Q}_{0,n})$  in every direction of  $\{q_i(dy_i; y^{i-1}, x^i) : i = 0, 1, \dots, n\}$  (due to nonstationarity).

**Theorem 4. (Gateaux Derivative)** Let  $\mathbb{I}_{\mu_{0,n}}(q_i : i = 0, 1, \dots, n) \triangleq \mathbb{I}_{X^n \rightarrow Y^n}(\mu_{0,n}, \vec{Q}_{0,n})$  be well defined for every  $\vec{Q}_{0,n} \in \mathcal{Q}^{C1}(\mathcal{Y}_{0,n}; \mathcal{X}_{0,n})$ . Then  $\{q_i(\cdot; \cdot, \cdot) : i = 0, 1, \dots, n\} \rightarrow \mathbb{I}_{\mu_{0,n}}(q_i(\cdot; \cdot, \cdot) : i = 0, 1, \dots, n)$  is Gateaux differentiable at every point in  $\mathcal{Q}(\mathcal{Y}_i; \mathcal{Y}_{0,i-1} \times \mathcal{X}_{0,i})$ , and the Gateaux derivative at the points  $q_i^*(\cdot; \cdot, \cdot)$  in each direction  $\delta q_i = q_i - q_i^*, i = 0, \dots, n$ , is

$$\begin{aligned} & \delta \mathbb{I}_{\mu_{0,n}}(q_i^*, q_i - q_i^* : i = 0, \dots, n) \\ &= \sum_{i=0}^n \int \log \left( \frac{q_i^*(dy_i; y^{i-1}, x^i)}{\nu_i^*(dy_i; y^{i-1})} \right) \frac{d}{d\epsilon} \vec{Q}_{0,i}^\epsilon(dy^i | x^i) \Big|_{\epsilon=0} \mu_{0,i}(dx^i) \log \int e^{s\rho_{0,i} - g_{i,n}} \nu_i^*(dy_i; y^{i-1}) \times \vec{Q}_{0,i-1}^*(dy^{i-1} | x^{i-1}) \mu_{0,i}(dx^i) \end{aligned}$$

where  $\vec{Q}_{0,i}^\epsilon \triangleq \otimes_{j=0}^i q_j^\epsilon(dy_j; y^{j-1}, x^j)$ ,  $q_j^\epsilon = q_j^* + \epsilon(q_j - q_j^*)$ ,  $j = 0, 1, \dots, i$ ,  $i = 0, 1, \dots, n$ ,

$$\begin{aligned} & \frac{d}{d\epsilon} \vec{Q}_{0,0}^\epsilon(dy_0; x_0) \Big|_{\epsilon=0} = \delta q_0(dy_0; x_0) \\ & \frac{d}{d\epsilon} \vec{Q}_{0,1}^\epsilon(dy^1 | x^1) \Big|_{\epsilon=0} = \delta q_0(dy_0; x_0) \otimes q_1^*(dy_1; y_0, x^1) \\ & \quad + q_0^*(dy_0; x_0) \otimes \delta q_1(dy_1; y_0, x^1) \\ & \vdots \\ & \frac{d}{d\epsilon} \vec{Q}_{0,i}^\epsilon(dy^i | x^i) \Big|_{\epsilon=0} = \delta q_0(dy_0; x_0) \otimes_{j=1}^i q_j^*(dy_j; y^{j-1}, x^j) \\ & \quad + q_0^*(dy_0; x_0) \delta q_1(dy_1 | y_0, x^1) \otimes_{j=2}^i q_j^*(dy_j; y^{j-1}, x^j) \\ & \quad \dots + \otimes_{j=0}^{i-1} q_j^*(dy_j; y^{j-1}, x^j) \otimes \delta q_i(dy_i; y^{i-1}, x^i). \end{aligned}$$

*Proof:* The derivation is lengthy, hence it is omitted. ■ The constrained problem defined by (8) can be reformulated using Lagrange multipliers (due to its convexity) by utilizing Lagrange Duality Theorem [10] to obtain

$$\begin{aligned} \vec{R}_{0,n}(D) &= \inf_{\vec{Q}_{0,n} \in \vec{\mathcal{Q}}_{0,n}(D)} \mathbb{I}_{X^n \rightarrow Y^n}(\mu_{0,n}, \vec{Q}_{0,n}) \\ &= \max_{s \leq 0} \inf_{\vec{Q}_{0,n} \in \vec{\mathcal{Q}}_{0,n}(D)} \left\{ \mathbb{I}_{X^n \rightarrow Y^n}(\mu_{0,n}, \vec{Q}_{0,n}) \right. \\ & \quad \left. - s \left( \ell_{d_{0,n}}(\vec{Q}_{0,n}) - (n+1)D \right) \right\} \end{aligned}$$

where  $s \in (\infty, 0]$  is the Lagrange multiplier.

Since  $q_i(dy; y^{i-1}, x^i) \in \mathcal{Q}(\mathcal{Y}_i; \mathcal{Y}_{0,i} \times \mathcal{X}_{0,i})$ , one should introduce another Lagrange multiplier to obtain an optimization problem without constraints. This process is involved; hence we state the final results.

**General Recursions of Optimal Non-stationary Reproduction Distribution.** The general recursions are the following. Define

$$\begin{aligned} g_{n,n}(x^n, y^n) &= 0 \\ g_{n-k,n}(x^{n-k}, y^{n-k}) &= - \int p_{n-k+1}(dx_{n-k+1}; x^{n-k}) \\ & \log \left( \int e^{s\rho_{0,n-k+1}(x^{n-k+1}, y^{n-k+1}) - g_{n-k+1,n}(x^{n-k+1}, y^{n-k+1})} \right. \\ & \quad \left. \times \nu_{n-k+1}^*(dy_{n-k+1}; y^{n-k}) \right), \quad k = 1, \dots, n. \end{aligned} \quad (16)$$

For  $k = 0, 1, \dots, n$ ,  $i \triangleq n - k$ , we get

$$\begin{aligned} q_n^*(dy_n; y^{n-1}, x^n) &= \frac{e^{s\rho_{0,n}(x^n, y^n)} \nu_n^*(dy_n; y^{n-1})}{\int_{\mathcal{Y}_n} e^{s\rho_{0,n}(x^n, y^n)} \nu_n^*(dy_n; y^{n-1})} \quad (17) \\ q_i^*(dy_i; y^{i-1}, x^i) &= \frac{e^{s\rho_{0,i}(x^i, y^i) - g_{i,n}(x^i, y^i)} \nu_i^*(dy_i; y^{i-1})}{\int_{\mathcal{Y}_i} e^{s\rho_{0,i}(x^i, y^i) - g_{i,n}(x^i, y^i)} \nu_i^*(dy_i; y^{i-1})}. \quad (18) \end{aligned}$$

The nonanticipative RDF is given by

$$\vec{R}_{0,n}(D) = s(n+1)D - \sum_{i=0}^n \int \left( \int g_{i,n} q_i^*(dy_i; y^{i-1}, x^i) + \log \int e^{s\rho_{0,i} - g_{i,n}} \nu_i^*(dy_i; y^{i-1}) \right) \times \vec{Q}_{0,i-1}^*(dy^{i-1} | x^{i-1}) \mu_{0,i}(dx^i)$$

**Discussion.** The above recursions illustrate the nonanticipation, since  $g_{i,n}(x^i, y^i)$ ,  $i = n - k$ ,  $k = 0, 1, \dots, n$ , appearing in the exponent of the reproduction distribution (18) integrate out future reproduction distributions. Note also that for the stationary case all reproduction conditional distributions are the same and hence,  $g_{i,n}(\cdot, \cdot) = 0$ , which implies  $q_n^*(\cdot; \cdot, \cdot)$  is given by (17). The above recursions are general, while depending on the assumptions imposed on the distortion function and source they can be simplified considerably.

## VI. CONCLUSION

In this paper, we derive an analytical closed form expression for the nonanticipative RDF for nonstationary processes, and we relate the definition of nonanticipative RDF to other works in the literature.

## REFERENCES

- [1] S. C. Tatikonda, "Control over communication constraints," Ph.D. dissertation, Mass. Inst. of Tech. (M.I.T.), Cambridge, MA, 2000.
- [2] A. K. Gorbunov and M. S. Pinsker, "Nonanticipatory and prognostic epsilon entropies and message generation rates," *Problems of Information Transmission*, vol. 9, no. 3, pp. 184–191, July–Sept. 1973.
- [3] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code, or not to code: Lossy source-channel communication revisited," *IEEE Transactions on Information Theory*, vol. 49, no. 5, pp. 1147–1158, May 2003.
- [4] P. A. Stavrou and C. D. Charalambous, "Nonanticipative rate distortion function and filtering theory: A weak convergence approach," *CoRR*, vol. abs/1212.6643v1, 2013. [Online]. Available: <http://arxiv.org/abs/1212.6643v1>.
- [5] J. L. Massey, "Causality, feedback and directed information," in *International Symposium on Information Theory and its Applications (ISITA '90)*, Nov. 27–30 1990, pp. 303–305.
- [6] M. S. Derpich and J. Østergaard, "Improved upper bounds to the causal quadratic rate-distortion function for gaussian stationary sources," *IEEE Transactions on Information Theory*, vol. 58, no. 5, pp. 3131–3152, May 2012.
- [7] C. D. Charalambous and P. A. Stavrou, "Directed information on abstract spaces: properties and extremum problems," in *IEEE International Symposium on Information Theory (ISIT)*, July 1–6 2012, pp. 518–522.
- [8] P. Dupuis and R. S. Ellis, *A Weak Convergence Approach to the Theory of Large Deviations*. John Wiley & Sons, Inc., New York, 1997.
- [9] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Englewood Cliffs, NJ: Prentice-Hall, 1971.
- [10] D. G. Luenberger, *Optimization by Vector Space Methods*. John Wiley & Sons, Inc., New York, 1969.